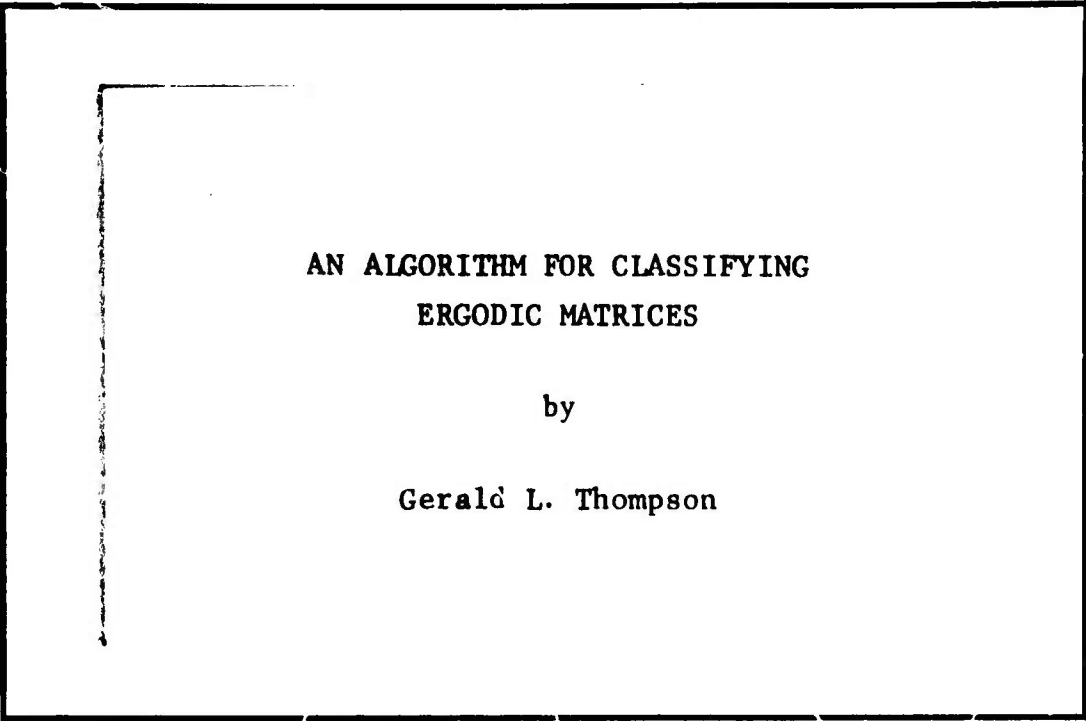


AD 641322

1-0



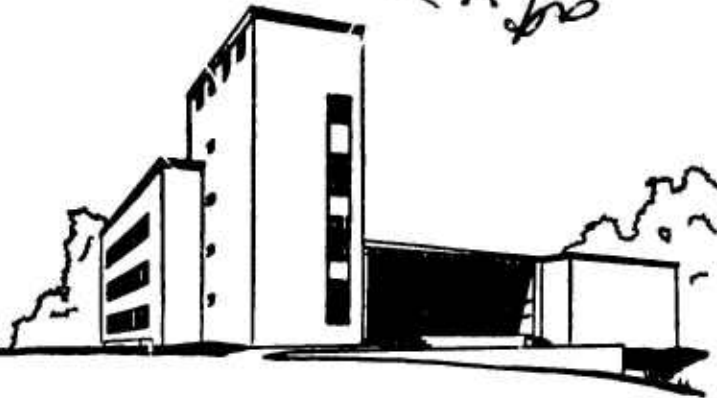
Carnegie Institute of Technology

Pittsburgh 13, Pennsylvania

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION			
Hardcopy	Microfiche	15	72
\$ 1.00	\$.50	pp	
1 ARCHIVE COPY			

NOV 7 1966

A 60



GRADUATE SCHOOL of INDUSTRIAL ADMINISTRATION

William Larimer Mellon, Founder

Management Sciences Research Report No. 83

AN ALGORITHM FOR CLASSIFYING
ERGODIC MATRICES

by

Gerald L. Thompson

September, 1966

Management Sciences Research Group
/ Graduate School of Industrial Administration
Carnegie Institute of Technology
Pittsburgh, Pennsylvania 15213

This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie Institute of Technology, under Contract NONR 760(24) NR 047-048 with the U. S. Office of Naval Research and under a Ford Foundation Grant. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

1. INTRODUCTION

An ergodic matrix is a nonnegative matrix with a special property to be defined in the next section. Such matrices occur in finite Markov chain theory, mathematical economics, sociology, and similar applications. In 1950 Wielandt [7] conjectured that an ergodic matrix was regular (to be defined) if and only if its $n^2 - 2n + 2$ power had all entries positive. The first published proof of this fact was by Rosenblatt [5] who also established many other facts concerning such matrices. Subsequently, two other proofs by Holladay and Varga [1] and Perkins [4] have appeared.

In this paper a new elementary completely independent proof is given that gives rise to some simple algorithms for determining whether an ergodic matrix is regular or cyclic. The work in carrying out these algorithms is considerably less (in the numerical analysis sense) than that of previous ones.

The author is indebted to Professor John G. Kemeny for advice during the preparation of the paper which led to substantial improvement of the algorithms.

2. PRELIMINARY RESULTS

Let A be a nonnegative $n \times m$ matrix. We shall denote the set of row and column indices by $S = \{1, \dots, n\}$ and call them states. Let $a_{ij}^{(t)}$ be the i, j th element of A^t . We say that state i can contact state j in t steps if $a_{ij}^{(t)} > 0$. The immediate successors of i are those states it can contact in one step, and the successors of i are those states it can contact in one or more steps. The immediate predecessors and predecessors of i are defined analogously. The matrix A is said to be ergodic if every

state as its successor, that is, every state can contact every other state. There are easy algorithms for checking for ergodicity, see e.g., [3], Chapter VII, Section 2.

Let $S_{i,t}$ be the set of states that i can contact in exactly t steps for $t \geq 1$. Define $S_{i,0} = \{i\}$. Let $T_{i,t} = S_{i,1} \cup \dots \cup S_{i,t}$ be the set of states i can contact in 1 through t steps.

LEMMA 1. (a) $S_{i,t+q}$ is uniquely determined by $S_{i,t}$ for $q \geq 1$.

(b) If $S_{i,t+m} = S_{i,t}$ for positive integers t and m then $S_{i,h+m} = S_{i,h}$ for all $h \geq t$.

(c) Every state can contact every other state in at most $n-1$ steps.

PROOF. (a) Note that $S_{i,1}$ consists of the column indices of the positive entries of the i th row of A , and hence is uniquely determined. Assume $S_{i,t-1}$ is uniquely determined by $S_{i,1}$. Then $S_{i,t}$ consists of the column indices of the positive entries of the rows whose indices are in $S_{i,t-1}$, which is a unique set. Hence $S_{i,t}$ is uniquely determined by $S_{i,t-1}$. So is $S_{i,t+q}$, and the statement follows.

(b) If $S_{i,t+m} = S_{i,t}$ for positive integers t and m , then it follows from (a) that $S_{i,h+1+m} = S_{i,h+1}$, hence (b) follows by finite induction.

(c) Consider the sequence

$$(1) \quad \{i\} = T_{i,0} \subseteq T_{i,1} \subseteq \dots \subseteq T_{i,t}.$$

Note that $T_{i,k} = T_{i,k+1}$ only if $S_{i,k+1} \subseteq T_{i,k}$, and (a) implies that the latter can hold only if $S_{i,h} \subseteq T_{i,k}$ for all $h \geq k$. Hence the sequence (1) is strictly increasing for a certain number of terms and then constant.

Because A is assumed to be ergodic, $T_{i,t} = S$ for t sufficiently large. However, the sequence (1) can be strictly increasing for at most $n-1$ steps, since at least one new state must be added each time and there are n states in all. Hence $T_{i,n-1} = S$, which is (c).

LEMMA 2. (a) For each i there is a smallest positive integer m for which $S_{i,t+m} = S_{i,t}$ for some t . Moreover $m \leq n$.

(b) For t sufficiently large $S = S_{i,t} \cup \dots \cup S_{i,t+m-1}$.

(c) The smallest m given by (a) is the same for every i .

(d) The sets $S_{i,t}, S_{i,t+1}, \dots, S_{i,t+m-1}$ partition S .

PROOF. (a) Since A is ergodic there is a smallest positive integer m such that i is a member of $S_{i,m}$. By Lemma 1(c), $m \leq n$. It follows that $S_{i,qm} \subseteq S_{i,(q+1)m}$ for $q = 0, 1, \dots$. Consider the sequence

$$(2) \quad S_{i,0} \subseteq S_{i,m} \subseteq S_{i,2m} \subseteq \dots \subseteq S_{i,(q+1)m}.$$

By Lemma 1(b) this sequence is strictly increasing for a few terms and then constant. Hence there is a smallest q such that $S_{i,qm} = S_{i,qm+m}$. Setting $t = qm$ we conclude that there is at least one m satisfying the condition. The set of possible m is finite and has a least member, which is the one whose existence is asserted in (a).

(b) Let m be as in (a) and let t be large enough that $S_{i,t+m} = S_{i,t}$. By Lemma 1(b) the sequence of $S_{i,h}$ for $h \geq t+m$ simply repeats the sequence $S_{i,t}, S_{i,t+1}, \dots, S_{i,t+m-1}$. Since A is ergodic S is contained in the union of these sets. And since S contains each one, it contains their union.

(c) (This proof is adapted from Kemeny and Snell [2], ppgs. 5-7.) Let N_{ij} be the set of times at which a message starting from i can reach j . In particular, N_{ii} is the set of times that a message can return to i .

The set N_{ii} is closed under addition, since if a and b are two such times then $a + b$ is also such a time. By a well known theorem a set of positive integers that is closed under addition contains all but a finite number of positive multiples of the greatest common divisor of the numbers in the set. Let d_i be the greatest common divisor of the numbers in N_{ii} .

In the next two paragraphs we let a and b be members of N_{ij} and c a member of N_{ji} .

Since a is a time that i can send to j and c a time that j can send to i , we see that $a+c$ belongs to N_{ii} ; hence $a+c \equiv 0 \pmod{d_i}$. It is also possible to send a message from i to j in time a , from j back to itself in time kd_j , and then from j to i in time c . Hence $a+kd_j+c \equiv 0 \pmod{d_i}$. Hence $kd_j = hd_i$ for some h . Since k can be any large prime number, d_j is divisible by d_i . By a symmetric argument d_i is divisible by d_j . Hence $d_i = d_j = d$, where d is the common divisor for all i .

Finally note that $a+c$ and $b+c$ both belong to N_{ii} , and hence are congruent mod d , hence $a \equiv b \pmod{d}$.

We know that $m_i \geq d$. But for k sufficiently large N_{ii} contains kd . Hence $m_i = d = m_j$ for all i and j .

(d) As we saw, $a \equiv b \pmod{d}$. If $S_{i,t+k} \cap S_{i,t+k'} \neq \emptyset$ for $k \neq k'$ and $0 \leq k, k' < m$, then there is some j such that i can send to j at times $t+k$ and $t+k'$ where $k'-k < d$, a contradiction. The fact that the union of $S_{i,t}, \dots, S_{i,t+m-1}$ is S is Lemma 2(b).

3. MAIN RESULTS

DEFINITION. If $m > 1$ then matrix A is cyclic with period m .

If $m = 1$, then A is regular.

THEOREM 1. (a) Every ergodic matrix is either cyclic or regular.

(b). A is regular if and only if there is a positive integer N such that $A^N > 0$.

(c) If $a_{ii} > 0$ for some i , then A is regular and $A^{2(n-1)} > 0$.

PROOF. (a) Obvious, since $m \geq 1$.

(b) Since for every i , $S_{i,t} = S_{i,t+1} = S$ for t sufficiently large, some power of A is positive. Conversely, if the N th power of A is positive, then $S_{i,N} = S_{i,N+1}$ so that $m = 1$.

(c) $S_{i,t} = T_{i,t}$ for every t . Hence $m = 1$. Since $S_{i,n-1} = T_{i,n-1} = S$ by Lemma 1(c), we see that $a_{i,k}^{(n-1)} > 0$ and $a_{k,i}^{(n-1)} > 0$ for all k . The definition of matrix multiplication now implies that $A^{2(n-1)} > 0$.

There exist matrices (see matrix C in Section 4) for which the bound in Theorem 1(c) is the best possible. However, for specific matrices better bounds are frequently possible. E.g., let k be the longest time it takes to send a message from any state j to i , and let k' be the longest time it takes to send from i to any other state. Each of these is less than or equal to $n-1$. It is easy to see that $A^{k+k'} > 0$.

THEOREM 2. If A is cyclic, then by reordering the states so that for some large t members of $S_{i,t}$ are listed first, $S_{i,t+1}$ second, etc., then A can be put in the following canonical form:

$$A = \begin{pmatrix} 0 & A_1 & 0 & \dots & 0 \\ 0 & 0 & A_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & A_{m-1} \\ A_m & 0 & 0 & \dots & \cdot \end{pmatrix}$$

where A_k is the submatrix with row indices in $S_{i,t+k-1}$ and column indices in $S_{i,t+k}$ for $k=1, \dots, m$.

PROOF. By Lemma 2(d), the sets $S_{i,t}, S_{i,t+1}, \dots, S_{i,t+m-1}$ partition S . By the definition of $S_{i,h}$, the set $S_{i,h+1}$ consists of the successors of $S_{i,h}$ and is uniquely determined, hence the entries outside the A_k 's are all zeros. Since A is ergodic every row and column of A , and hence of A_k must have a nonzero entry, therefore no further decomposition of A is possible.

THEOREM 3. If A is an $n \times n$ nonnegative ergodic matrix then either $A^{n^2-2n+2} > 0$ and A is regular or else A is cyclic.

PROOF. Suppose A is regular. Then $S_{i,t} = S$ for every i and for t sufficiently large. For each state i there exists a minimum t such that $S_{i,t} = S$. Let i_0 be a state whose minimal t is maximal. We must show that $t \leq n^2-2n+2$. Denote $S_{i_0,t}$ by S_t from now on.

Let k be the smallest index such that $i_0 \in S_k$. We know that $1 \leq k \leq n$. If $k=1$, then by Theorem 1(c) $A^{2(n-1)} > 0$. Since $(n^2-2n+2)-2(n-1) = (n-2)^2 \geq 0$, the theorem is true in this case.

Suppose now that $1 < k \leq n$. Choose q to be the largest positive integer so that $S_{qk} \neq S_{(q+1)k}$. Consider the sequence

$$\{i_0\} = S_0, S_1, \dots, S_k, \dots, S_{qk}, S_{qk+1}, S_{qk+2}, \dots, S_{(q+1)k}, \dots$$

Let j_0 be any element different from i_0 in S_1 ; obviously such an element exists. Then $j_0 \in S_{qk+1}$ for all $q \geq 1$. Consider the sequence,

$$S_0 \subset S_k \subset S_{2k} \subset \dots \subset S_{qk}$$

$$S_1 \subset S_{k+1} \subset S_{2k+1} \subset \dots \subset S_{qk+1}$$

The containing relationships in the second sequence are a consequence of those in the first sequence. Both these sequences are strictly increasing by Lemma 1(b). Hence we can choose the following sequence of elements:

$$i_v \in S_{vk} - S_{(v-1)k}$$

$$j_v \in S_{vk+1} - S_{(v-1)k+1}$$

for $v=1, \dots, q$. Because $i_0 \neq j_0$ this can be done at most $n-2$ times, so that $q \leq n-2$. If $q < n-2$ then $q+1 \leq n-2$ and $S_{(q+1)k} = S$. But $(q+1)k \leq (n-2)n < n^2 - 2n + 2$, and the theorem is true.

Hence suppose that $q=n-2$. Then

$$S_{qk} = \{i_0, i_1, \dots, i_{n-2}\} = S - \{j_0\}$$

$$S_{qk+1} = \{j_0, j_1, \dots, j_{n-2}\} = S - \{i_0\}$$

Now i_0 will be in S_{qk+2} since one of its predecessors is in S_{qk+1} .

Also S_{qk+2} will contain $S_{qk} \cap S_{qk+1} = S - \{i_0, j_0\}$ since all these elements have appeared in two adjacent sets. Hence $S_{qk+2} \supseteq S_{qk}$. If $S_{qk+2} = S_{qk} = S - \{j_0\}$, then by Lemma 1(b) A is cyclic with period 2. Hence $S_{qk+2} = S$. Now

$$qk+2 \leq (n-2)n + 2 = n^2 - 2n + 2.$$

Since i_0 was a state with largest minimal t , $A^{n^2 - 2n + 2} > 0$.

4. THE ALGORITHM

Theorem 3 can, of course, be used to check on regularity of A by computing powers of A . Since we can square A each time, we need only square k times with k being the smallest integer such that $2^k > n^2 - 2n + 2$. However, squaring a matrix requires n^3 multiplications so that this represents considerable work.

A much more efficient algorithm (in the sense of requiring fewer numerical operations) can be based on Lemma 2(c) and (d), and the obvious fact that if two successive S_j 's have a common element then the matrix is regular, since the partition of Lemma 2(d) can then have but one set. No numerical operations are involved--merely lookups to see for each state what its successors are. The algorithm may be described as follows:

0. Let i be any row index (a good start is to let it be a row with maximum number of positive entries). Let $S_0 = U = \{i\}$; let $j = t = 0$; let $m = \infty$.
1. Let $U = S_j$
2. Let $j = j+1$, $t = t+1$.
3. Let $S_j = \{\text{the immediate successors of elements of } U\}$.
4. If $S_j \cap U \neq \emptyset$ go to 11.
5. If i is not in S_j go to 1.
6. If $t = m$ go to 1.
7. Let $t = m$.
8. If $m = 1$ go to 11.
9. If $j = n^2 - 2n + 2$ go to 12.
10. Let $t = 0$; go to 1.
11. Stop. The matrix is regular.

12. Stop. The matrix is cyclic with period m . The partition of the states is given by $S_j, S_{j+1}, \dots, S_{j+m-1}$.

There is an additional instruction that can be added to the algorithm to speed it up, at the expense of requiring more memory space. The instruction should be inserted between 4 and 5, and is

4a. If $S_j = S_{j-m}$ go to 12.

Note that to implement this instruction all the sets between S_{j-m} and S_j must either be saved, or recomputed when necessary. A more economical change that requires at most $m-1$ additional iterations of the program is to number the additional instruction 5a, and insert it between 5 and 6.

To illustrate the use of these algorithms consider the matrices,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The complete calculations that show that A is regular and B cyclic are given below, using the 4a instruction in the algorithm.

	S_0	S_1	S_2	S_3	S_4	S_5
For A	{6}	{1,2}	{2,3}			
For B	{6}	{1,5}	{2,6}	{1,3,5}	{2,4,6}	{1,3,5}

Note that the algorithm is such that it can be done by hand, even for fairly large matrices. Since for $n=6$, $n^2-2n+2 = 26$, we see that these computations are considerably simpler than raising the matrices to the 26th

power. It can be shown that A and B are such that their ergodic properties are not evident before the 26th power. Matrix C has the property that $C^9 > 0$ and $C^{10} > 0$, showing that the bound in Theorem 1(c) is the best possible.

In [6] the author has published a computer program (in the language BASIC) that implements the algorithm using the 5a instruction.

BIBLIOGRAPHY

- [1] Holladay, J. C., and P. S. Varga, "On Powers of Nonnegative Matrices," Proc. Amer. Math. Soc., Vol. IX (1958) pp. 631-634.
- [2] Kemeny, J. G., and J. L. Snell, Finite Markov Chains. Princeton, N. J., Van Nostrand, 1960.
- [3] Kemeny, J. G., J. L. Snell, and G. L. Thompson, Introduction to Finite Mathematics, Second Edition, Englewood Cliffs, N.J., Prentice-Hall, 1966, pp. 394-406.
- [4] Perkins, P., "A Theorem on Regular Matrices," Pac. J. of Math., (1961) p. 1529-1533.
- [5] Rosenblatt, D., "On Graphs and Asymptotic Forms of Finite Boolean Relation Matrices and Stochastic Matrices," Nav. Log. Res. Quar. IV (1957) pp.151-167.
- [6] Thompson, G. L., Teaching Constructive Linear Algebra with the Aid of a Computer, mimeographed teaching supplement available through the Department of Mathematics, Dartmouth College.
- [7] Wielandt, H., "Unzerlegbar, nicht negativen Matrizen," Math. Zeit., 52 (1950) pp. 642-648.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Graduate School of Industrial Administration Carnegie Institute of Technology		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP Not applicable	
3. REPORT TITLE AN ALGORITHM FOR CLASSIFYING ERGODIC MATRICES			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report, September, 1966			
5. AUTHOR(S) (Last name, first name, initial) Thompson, Gerald L.			
6. REPORT DATE September, 1966		7a. TOTAL NO. OF PAGES 11	7b. NO. OF REFS 7
8a. CONTRACT OR GRANT NO. NONR 760(24)		8a. ORIGINATOR'S REPORT NUMBER(S) Management Sciences Research Report No. 83	
b. PROJECT NO. NR 047-048		8b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None	
c.			
d.			
10. AVAILABILITY/LIMITATION NOTICES Distribution of this document is unlimited			
11. SUPPLEMENTARY NOTES None		12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D. C. 20360	
13. ABSTRACT An algorithm is derived for determining whether a given ergodic matrix is regular or cyclic, and for putting it in canonical form in the latter case. It requires very little numerical computation and is suitable for hand use even for fairly large matrices. For large matrices, the corresponding program is considerably faster than previous ones. Paper also contains elementary independent proofs of several known results on ergodic matrices.			

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Canonical form of a nonnegative matrix Ergodic matrix Markov chain graph theory sociometric matrix						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parentheses immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.